**Laboratory Assignments**

**Subject: Machine Learning**

---

1. Perform linear regression to predict

   a. $CO_2$ Emission **Dataset: fuel_consumption_dataset.csv**

   b. The selling price of a used car. **Dataset: used_cars_dataset.csv**

   Evaluate the quality of the models by computing relevant performance metrics, including the $R^2$ value. Generate and display a plot that compares the actual values to the predicted values (Actual vs Predicted) for both tasks.

2. Perform linear regression with L1 (Lasso) and L2 (Ridge) regularization to predict the price of a House. Use hyper-parameter tuning for the best result. Evaluate the accuracy of the models by computing relevant performance metrics, including the $R^2$ value. Generate and display a plot that compares the actual values to the predicted values (Actual vs Predicted) for both tasks.

   **Dataset: housing_price_dataset.csv**

3. Perform linear regression with one feature using gradient descent (without using library function) to predict the salary of an employee based on the feature Years Experience. Use hyper-parameter tuning for the best result. Plot the hypothesis function and the data points after each epoch. Evaluate the accuracy of the models by computing relevant performance metrics, including the $R^2$ value.

   **Dataset: salary_dataset.csv**

4. Perform a non-linear regression to predict China's GDP from 1960 to 2014 from given features. Evaluate the quality of the model by computing relevant performance metrics, including the $R^2$ value. Generate and display a plot that compares the actual values to the predicted values (Actual vs Predicted) for both tasks.

   **Dataset: china_gdp.csv**

5. Perform logistic regression to classify if a patient has a benign tumor or malignant tumor (cancer) based on the features provided. Generate the confusion matrix and evaluate the quality of the model by computing relevant performance metrics including Precision, Recall, accuracy, F1-Score etc. Plot the ROC curve and calculate AUC.

   **Dataset: samples_cancer.csv**

6. Using KNN algorithm, predict which category a customer belongs to on the basis of the data provided by a telecommunications firm. Find the accuracy of the KNN algorithm in predicting the category of a customer.

   **Dataset: teleCust.csv**

7. Using the Decision Tree algorithm, predict which drug among drug X, drug Y and drug C should be given to a patient. Find the accuracy of the decision tree in predicting the correct drug for the patient.

   **Dataset: drug.csv**

8. Using Naive Bayes algorithms, predict if a person is diabetic or not, based on the features provided. Find the accuracy and F1-Scores of both algorithms.

   **Dataset: pima-indians-diabetes.data.csv**

9. Using SVM algorithm, predict if a patient has a benign tumor or malignant tumor (cancer) based on the features provided. Use the following kernel for the SVM algorithm:

   a) Linear b) Polynomial c) RBF d) Sigmoid

   Find the following metrics for each of the SVM algorithms:

   1) Accuracy 2) Recall 3) Precision 4) F1-Score

   5) Jaccard Score 6) Error rates 7) Confusion Matrix

   Compare all four SVM models using an ROC curve.

   **Dataset: samples_cancer.csv**

10. Applying SVM, Naive Bayes, Decision tree and KNN to predict diabetes based on features set. Compare the four classification algorithms with performance metrics such as accuracy, recall, precision, F1- score. Also, design the heat map confusion matrix for the above algorithms and construct ROC curve for comparison.

    **Dataset: pima-indians-diabetes.data.csv**

11. Design and train a network of perceptrons that computes the functionality of XOR.

12. Perform a Multilayer perceptron neural network to classify flower type. Utilize number of hidden layers, 5 and 200 to 400 iterations with a learning rate. Try with different loss functions/ activation functions such as MSE, Cross entropy, sigmoid, tanh, ReLU along with different optimizers GD, SGD, Adam. Illustrate the result with performance metrics and observe Weight, Loss curve and accuracy curve.

    **Dataset: Iris dataset**

13. Perform k- means clustering algorithm for customer segmentation from given features. Utilize Euclidean distance and Manhattan distance for this problem. Also, plot in terms of 2D and 3D clusters this problem.

    **Dataset: Customer segmentation dataset**

14. Perform hierarchical clustering such as Agglomerative algorithm and Divisive algorithm to group several vehicles. Utilize single, complete, and average linkage to define the cluster. Also draw the dendrogram for this problem.

    **Dataset: Vehicle dataset**

15. Perform the DBSCAN algorithm for weather station clustering. Utilize proper data cleaning and feature selection. Also, plot all outliers of the cluster label.

**Dataset: Weather Station dataset**